

**This Page Is Inserted by IFW Operations  
and is not a part of the Official Record**

## **BEST AVAILABLE IMAGES**

**Defective images within this document are accurate representations of the original documents submitted by the applicant.**

**Defects in the images may include (but are not limited to):**

- **BLACK BORDERS**
- **TEXT CUT OFF AT TOP, BOTTOM OR SIDES**
- **FADED TEXT**
- **ILLEGIBLE TEXT**
- **SKEWED/SLANTED IMAGES**
- **COLORLED PHOTOS**
- **BLACK OR VERY BLACK AND WHITE DARK PHOTOS**
- **GRAY SCALE DOCUMENTS**

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning documents *will not* correct images,  
please do not report the images to the  
Image Problem Mailbox.**

# PATENT ABSTRACTS OF JAPAN

(11)Publication number : 08-339446

(43)Date of publication of application : 24.12.1996

(51)Int.Cl.

G06T 7/60  
G06F 3/16  
G06F 13/00  
G06T 1/00  
G10L 3/00  
G10L 3/00  
G10L 9/00

(21)Application number : 07-143511

(71)Applicant : SHARP CORP  
GIJUTSU KENKYU KUMIAI SHINJOHO SHIYORI  
KAIHATSU KIKO

(22)Date of filing : 09.06.1995

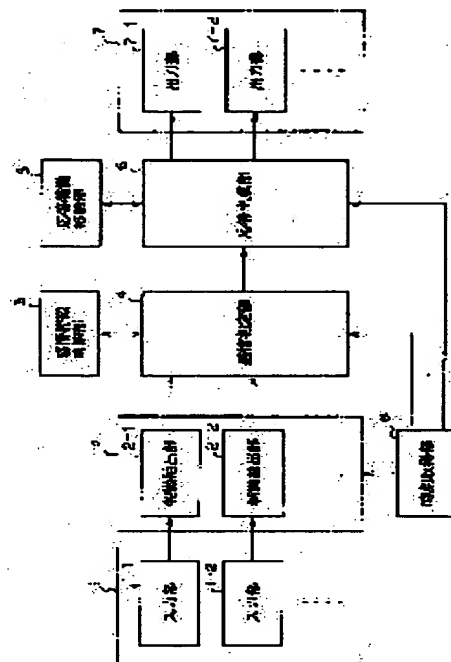
(72)Inventor : WATANUKI KEIKO

## (54) INTERACTIVE SYSTEM

### (57)Abstract:

**PURPOSE:** To provide an interactive system between a user (human) and a computer that a user feels familiar by detecting diverse feelings that the user has and outputting information from the computer side.

**CONSTITUTION:** This system consists of plural input parts 1 (1-1, 1-2...) which react to the operation and behavior of the user, feature extraction parts 2 (2-1, 2-2...) which extract features of signals inputted from the input parts 1, a feeling decision part 4 which decides the feelings of the user from plural signal features extracted by the feature extraction parts 2, a response generation part 6 which generates the response contents of the computer on the basis of the feelings decided by the feeling decision part 4, and output parts 7 (7-1, 7-2...) for the response contents. The response contents are transmitted to the user by the output parts 7.



## LEGAL STATUS

[Date of request for examination] 18.12.1998

[Date of sending the examiner's decision of rejection] 07.11.2000

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

(19)日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11)特許出願公開番号

特開平8-339446

(43)公開日 平成8年(1996)12月24日

(51)Int.Cl. <sup>9</sup>	識別記号	序内整理番号	F I	技術表示箇所
G 0 6 T 7/60		9061-5H	G 0 6 F 15/70	3 5 0 Z
G 0 6 F 3/16	3 2 0	9172-5E	3/16	3 2 0 A
13/00	3 5 3	7368-5E	13/00	3 5 3 S
G 0 6 T 1/00			G 1 0 L 3/00	5 3 1 N
G 1 0 L 3/00	5 3 1			5 7 1 G

審査請求 未請求 請求項の数 3 O L (全 8 頁) 最終頁に続く

(21)出願番号 特願平7-143511

(22)出願日 平成7年(1995)6月9日

(71)出願人 000005049

シャープ株式会社

大阪府大阪市阿倍野区長池町22番22号

(71)出願人 593162453

技術研究組合新情報処理開発機構

東京都千代田区東神田2-5-12 龍角散ビル8階

(72)発明者 綿貫 啓子

大阪府大阪市阿倍野区長池町22番22号 シャープ株式会社内

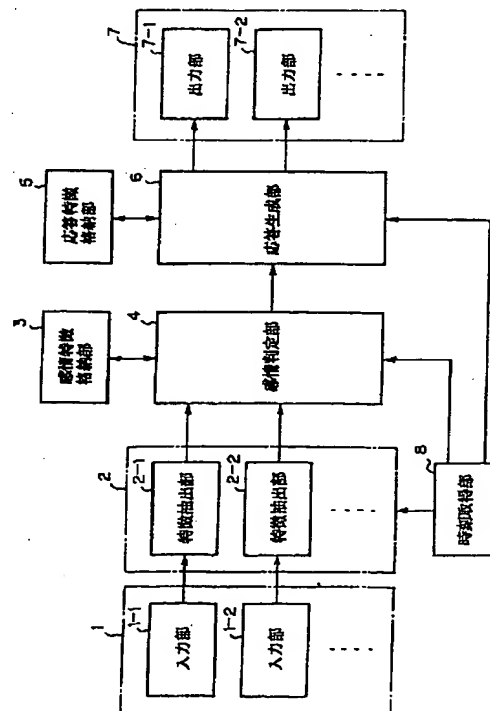
(74)代理人 弁理士 高野 明近 (外1名)

## (54)【発明の名称】 対話装置

## (57)【要約】

【目的】 ユーザ（人間）とコンピュータの対話装置において、ユーザの多様な感情を検出し、この感情に応じて、コンピュータ側から情報を出力することにより、親しみの持てる対話装置を提供すること。

【構成】 ユーザの動作や行動に応じる複数の入力部1（1-1, 1-2, …）と、入力部1から入力された信号の特徴を抽出する特徴抽出部2（2-1, 2-2, …）と、特徴抽出部2により抽出された複数の信号特徴からユーザの感情を判定する感情判定部4と、感情判定部4により判定された感情に基づき、コンピュータの応答内容を生成する応答生成部と、応答内容の出力部7（7-1, 7-2, …）から構成される。応答内容は、出力部7によりユーザに伝えられる。



(2)

1

## 【特許請求の範囲】

【請求項1】 ユーザ（人間）とコンピュータが音声あるいは表情などを通じて対話する対話装置において、前記ユーザの行動或いは動作に応じる複数の入力手段と、該入力手段から入力された信号の特徴を抽出する特徴抽出手段と、該特徴抽出手段により抽出された複数の信号特徴から前記ユーザの感情を判定する感情判定手段と、該感情判定手段により判定された感情に基づき、前記コンピュータの応答内容を生成する応答生成手段とから構成されることを特徴とする対話装置。

【請求項2】 前記感情判定手段は、前記複数の信号特徴として前記ユーザの音声の高さと視線の方向を抽出し、それらからユーザの感情を判定することを特徴とする請求項1記載の対話装置。

【請求項3】 感情の履歴を蓄積する履歴格納手段を更に備えたことを特徴とする請求項1または2記載の対話装置。

## 【発明の詳細な説明】

## 【0001】

【産業上の利用分野】本発明は、ユーザ（人間）とコンピュータとが対話する対話装置に関し、より詳細には、音声或いは表情などを通じて対話を行うためのものに関する。

## 【0002】

【従来の技術】従来、人間とコンピュータの間のインタフェースとしては、キーボードや手書き文字認識、音声認識などが知られている。しかし、これらの手段によってコンピュータ側に入力される情報は、言語に変換して入力されるものであり、入力を行う人間の感情を言語以外の情報として扱う手段を有するものではなかった。一方、特開平5-12023号公報には、音声認識を利用して使用者の感情を認識する装置が開示されている。また、特開平6-67601号公報には、手話使用者の表情を認識し、話者の感情を含んだ自然文を出力する装置が開示されている。さらに、特開平5-100667号公報には、演奏者の動きを検出して、演奏者の感情にマッチした楽音制御をする装置が開示されている。

## 【0003】

【発明が解決しようとする課題】人間が働きかけることを要する上述の従来例の装置と同様に、コンピュータとの対話においても、人間は気分が乗ってきたり、あるいはいらいらしたり、退屈したりと様々な感情を持つ。このような感情に対応すべく、特開平5-12023号公報では、感情を音声から抽出しようとするものであり、特開平6-67601号公報では、手話に伴う表情から捉えようとするものであり、また、特開平5-100667号公報では、演奏者の腕の曲げ押し等の体の動きから感情を検出しようとするものであるが、本来、人間の感情は、音声のみ、表情のみ、あるいは動きのみというように、シングルモードに現われるのではなく、音声や

2

表情、身振りなどと同時に、あるいは、相補的に現われるものであるから、従来例の手段は、必ずしも満足できるものではない。さらに、ユーザ（人間）の感情を検出しても、上述の従来例における装置の応答においても同様のことがいえるが、従来のコンピュータとの対話においては、コンピュータ側からの応答内容および応答の仕方がユーザの感情にかかわらず一定で、面白みのないものであった。本発明は、上述の課題を解決するためになされたもので、ユーザ（人間）とコンピュータの対話装置において、ユーザの多様な感情を検出するとともに、さらにこの感情に応じて、コンピュータ側から情報を出力することにより、親しみの持てる対話装置を提供することをその目的とする。

## 【0004】

【課題を解決するための手段】本発明は、上述の課題を解決するために、（1）ユーザ（人間）とコンピュータが音声あるいは表情などを通じて対話する対話装置において、前記ユーザの行動或いは動作に応じる複数の入力手段と、該入力手段から入力された信号の特徴を抽出する特徴抽出手段と、該特徴抽出手段により抽出された複数の信号特徴から前記ユーザの感情を判定する感情判定手段と、該感情判定手段により判定された感情に基づき、前記コンピュータの応答内容を生成する応答生成手段とから構成されること、或いは、（2）前記（1）において、前記感情判定手段は、前記複数の信号特徴として前記ユーザの音声の高さと視線の方向を抽出し、それらからユーザの感情を判定すること、或いは、（3）前記（1）又は（2）において、感情の履歴を蓄積する履歴格納手段を更に備えたことを特徴とするものを構成する。

## 【0005】

【作用】請求項1の対話装置においては、入力手段によりユーザの行動或いは動作に対応して発生する複数の信号から信号抽出手段によりユーザの複数の信号特徴が抽出される。そして、これら複数の信号特徴を統合的に扱い、感情判定手段によりユーザの感情を判定することができる。また、判定された感情に基づき、応答生成手段によりコンピュータからの応答が決定される。これにより、ユーザの感情に応じてコンピュータ側からの応答を制御することができるので、より親しみの持てる対話装置を提供することができる。請求項2の対話装置においては、音声の高さと視線の方向とからユーザの感情が判定される。これにより、より間違いの少ないユーザの感情を判定できる。請求項3の対話装置においては、履歴格納手段によりユーザの感情の履歴が蓄積される。これにより、ユーザの感情の変化を記録することができるようになり、ユーザの感情の変化に応じた感情判定ができるようになるとともに、ユーザの感情の変化に応じたコンピュータの応答の制御ができるようになるので、より満足のできる対話装置が得られる。

3

【0006】

【実施例】図1は、本発明の対話装置の実施例を示すブロック図である。図1において、1は、入力部、2は、入力部から得られる信号の特徴を抽出する特徴抽出部である。3は、感情を判定するためのデータをあらかじめ格納しておく感情特徴格納部であり、4は、感情特徴格納部3のデータを基に、ユーザの行動或いは動作から得られる信号の特徴からユーザの感情を判定する感情判定部である。5は、ユーザの感情に応じてコンピュータが出力すべきデータをあらかじめ格納しておく応答特徴格納部であり、6は、応答特徴格納部7のデータを基に、コンピュータの応答内容を生成する応答生成部である。7は、該応答生成部6により生成されたデータを出力する出力部である。8は、現在時刻を得るための時刻取得部である。

【0007】次に、本実施例の動作に関して説明する。入力部1は、例えばカメラやマイク、動きセンサ、あるいは心電計など、複数の入力部1-1、1-2、…を備えることができ、ユーザの行動或いは動作に対応して発生する複数の信号が取り込まれる。特徴抽出部2で抽出される特徴としては、例えば、音声の高低（以下ピッチという）、音声の大きさ、発話の速度、ポーズの長さ、表情、顔の向き、口の大きさや形、視線の方向、身振り、手振り、頭の動き、心拍数などが考えられ、そのための複数の特徴抽出部2-1、2-2、…を備える。また、出力部7は、例えば、スピーカやディスプレイ、触覚装置など、複数の出力部7-1、7-2、…を備えることができる。

【0008】以下では、入力部1-1として音声を入力するための音声入力部を、入力部1-2としてユーザの顔画像を入力するための顔画像入力部を、また、特徴抽出部2-1としてユーザが発声する音声の高さを抽出するピッチ抽出部を、特徴抽出部2-2としてユーザの視線方向を検出し、コンピュータに視線を向けているかどうか（アイコンタクト）を判定する視線検出部を、さらに、出力部7-1としてCGによる疑似人間を表示する表示部、および出力部7-2として合成音声出力する音声出力部として、本発明の実施例が示されているので、その動作を説明する。マイク等の入力部1-1によって装置に取り込まれた音声信号は、特徴抽出部2-1でA/D変換され、あらかじめ決められた処理単位（フレーム：1フレームは1/30秒）毎に平均ピッチ[Hz]が求められ、フレーム毎の平均ピッチ変化量[%]が感情判定部4に送出される。カメラ等の入力部1-2によって装置に取り込まれた視線の画像は、特徴抽出部2-2でフレーム毎にアイコンタクトの時間長[sec]が求められ、フレーム毎のアイコンタクト時間長の変化量[%]が感情判定部4に送出される。

【0009】図2は、特徴抽出部2-1で抽出された平均ピッチ[Hz]の例を示す図である。また、図3は、時系

(3)

4

列にとったフレーム毎の平均ピッチ変化量[%]の例を示す図である。ここで、(+)数値はピッチが先行フレームより上がっていることを意味し、また、(-)数値は下がっていることを意味する。図4は、特徴抽出部2-2で検出されるアイコンタクトの時間長[sec]の例を示す図である。また、図5は、時系列にとったフレーム毎のアイコンタクト変化量[%]の例を示す図である。ここで、(+)数値はアイコンタクトの時間長が先行フレームより長くなっていることを意味し、また、(-)数値は短くなっていることを意味する。なお、ここでは、平均ピッチの変化量をピッチ特徴、およびアイコンタクトの時間長の変化量を視線特徴としたが、最高ピッチやアイコンタクトの回数などをそれぞれピッチ特徴、視線特徴としてもよい。

【0010】感情判定部4では、入力されたユーザのピッチ特徴および視線特徴を、フレーム毎に感情特徴格納部3のデータを参照して、該フレーム毎のユーザの感情が判定される。表1は、感情特徴格納部3のデータの例を示す表である。この表には、平均ピッチの変化量[%]とアイコンタクトの時間長の変化量[%]から判定されるユーザの感情として両者の関係が示されている。

【0011】

【表1】

入力情報		感情判定
ピッチ変化量	アイコンタクト変化量	
ピッチ>+20%	アイコンタクト>+40%	楽しい
ピッチ<-20%	アイコンタクト<-40%	退屈
ピッチ>+20%	アイコンタクト<-40%	イライラ
ピッチ<-20%	アイコンタクト>+40%	怒っている

【0012】図6は、感情判定部4での時系列にとったフレーム毎の処理の例を示す図である。ここでは、例えば、ピッチ変化量が+30[%]およびアイコンタクト変化量が+45[%]と検出され、ユーザの感情が「楽しい」と判定されている。

【0013】感情判定部4で判定された感情は、応答生成部6に送出される。該応答生成部6では、フレーム毎に応答特徴格納部5のデータを参照して、出力すべき音声情報および顔画像情報がそれぞれ出力部7-2と出力部7-1に送出される。表2は、応答特徴格納部5のデータの例を示す表である。この表には、ユーザの感情に応じてコンピュータによる応答をピッチパターンおよびCG顔画像で指定するようにするための両者の対応関係が示されている。もちろん、音声の大きさや発話の速度を指定したり、また、顔だけでなく、身振りも加えるようにしてもよい。

【表2】

ユーザの感情	コンピュータによる応答	
	音声ビッチ情報	顔面表情情報
楽しい	ビッチバシ2	笑顔
退屈	ビッチバシ2	困感顔
イライラ	ビッチバシ3	しかめ面
怒っている	ビッチバシ4	悲しい顔

【0015】図7は、感情判定部4での処理に応答生成部6での時系列にといったフレーム毎の処理の例を示す図である。ここでは、例えば、ユーザの「楽しい」という感情判定に対して、コンピュータからビッチバシ2の音声で笑顔のCG顔面像を出力するように処理している。図8は、応答生成部6で指定されるビッチバシの例を示す図で、また、図9は、CG顔面像の例を示す図である。

【0016】次に、本願のほかの発明の実施例を説明する。図10は、この実施例の装置構成を示すブロック図であり、図示のように、先の本発明の実施例の構成に、ユーザの感情の履歴を蓄積する履歴格納部9が付加されている。以下に、この実施例でユーザの感情の履歴を処理する動作について説明する。まず、先の実施例と同様の手順によって、感情判定部4で判定されたユーザの感情をフレーム毎に履歴格納部9に蓄積する。人間の感情は変化し、その感情変化には、たとえば、「楽しい」から「ふつう」の感情、「イライラ」の感情から「怒っている」感情というように、一定の規則制があると考えられる。そこで、感情判定部4では、該当フレームでのユーザのビッチ特徴および視線特徴と、さらに前フレームの感情の履歴を参照して、該当フレームのユーザの感情が判定される。図11は、ユーザの感情の履歴情報を利用した、感情判定部4での時系列にといったフレーム毎の処理の例を示す図である。ここでは、該当フレームで、ビッチ変化量が-30 [%] およびアイコンタクト変化量が+45 [%] と検出され、かつ、前フレームの感情履歴「イライラ」を参照して、ユーザの感情が「怒っている」と判定されている。感情判定部4で判定された感情は、応答生成部6に送出される。コンピュータとの対話において、ユーザの感情に応じて、コンピュータからの応答内容や応答の仕方が変化するようになれば、対話がより楽しいものになると考えられる。そこで、応答生成部6では、感情判定部4で判定された該当フレームの感情と履歴格納部9に蓄積された前フレームのデータを基に、該当フレームでのコンピュータからの応答が決定されて、出力部7-1、7-2...に送出される。図12は、ユーザの感情の履歴情報を利用した、応答生成部6での

(4)

コンピュータによる応答（アナログ制御）	
音声ビッチ情報	顔面表情情報
平均ビッチ＝ 「感情度」×K1	顔面像の口の形＝ 「感情度」×K2

【表3】

【0017】

る。なお、表中のK1、K2は係数である。

図10は、この実施例の装置構成を示すブロック図であり、図示のように、先の本発明の実施例の構成に、ユーザの感情の履歴を蓄積する履歴格納部9が付加されている。以下に、この実施例でユーザの感情の履歴を処理する動作について説明する。まず、先の実施例と同様の手順によって、感情判定部4で判定されたユーザの感情をフレーム毎に履歴格納部9に蓄積する。人間の感情は変化し、その感情変化には、たとえば、「楽しい」から「ふつう」の感情、「イライラ」の感情から「怒っている」感情というように、一定の規則制があると考えられる。そこで、感情判定部4では、該当フレームでのユーザのビッチ特徴および視線特徴と、さらに前フレームの感情の履歴を参照して、該当フレームのユーザの感情が判定される。図11は、ユーザの感情の履歴情報を利用した、感情判定部4での時系列にといったフレーム毎の処理の例を示す図である。ここでは、該当フレームで、ビッチ変化量が-30 [%] およびアイコンタクト変化量が+45 [%] と検出され、かつ、前フレームの感情履歴「イライラ」を参照して、ユーザの感情が「怒っている」と判定されている。感情判定部4で判定された感情は、応答生成部6に送出される。コンピュータとの対話において、ユーザの感情に応じて、コンピュータからの応答内容や応答の仕方が変化するようになれば、対話がより楽しいものになると考えられる。そこで、応答生成部6では、感情判定部4で判定された該当フレームの感情と履歴格納部9に蓄積された前フレームのデータを基に、該当フレームでのコンピュータからの応答が決定されて、出力部7-1、7-2...に送出される。図12は、ユーザの感情の履歴情報を利用した、応答生成部6での

【0018】

【発明の効果】人間とコンピュータが音声あるいは表情などを通じて対話する対話装置において、請求項1の対話装置においては、ユーザの行動に対応して発生する複数の信号特徴からユーザの感情を判定することができる。とともに、ユーザの感情に応じてコンピュータ側から応答するよう制御することができる。したがって、より親しみの持てる対話装置を提供できる。請求項2の対話装置においては、ユーザの音声の高さ（ビッチ）と視線の方向（アイコンタクト）とからユーザの感情を判定するので、より間違いの少ない判定が可能となる。請求項3の対話装置においては、ユーザの感情の変化に応じた感情判定ができるようになる。ユーザの感情の変化に応じてコンピュータの応答の制御ができるようになるので、対話装置として、より満足できるものが得られる。

【図面の簡単な説明】

【図1】本発明の対話装置の実施例を示すブロック図である。

【図2】本発明の実施例の特徴抽出部で抽出された平均

(5)

7

ピッチ [Hz] の例を示す図である。

【図3】本発明の実施例の特徴抽出部で抽出された平均ピッチの変化量 [%] の例を示す図である。

【図4】本発明の実施例の特徴抽出部で抽出されたアイコンタクト時間長 [sec] の例を示す図である。

【図5】本発明の実施例の特徴抽出部で抽出されたアイコンタクト時間長の変化量 [%] の例を示す図である。

【図6】本発明の実施例の感情判定部での処理の例を示す図である。

【図7】本発明の実施例の応答生成部での処理の例を示す図である。

【図8】本発明の実施例の応答生成部でのピッチパタンの例を示す図である。

【図9】本発明の実施例の応答生成部でのCG顔画像の

8

例を示す図である。

【図10】本発明の他の実施例の概略構成ブロック図である。

【図11】本発明の他の実施例のユーザの履歴を利用した感情判定部での処理の例を示す図である。

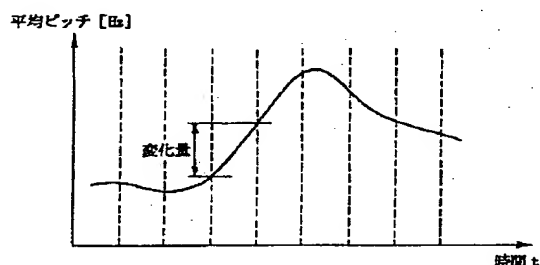
【図12】本発明の他の実施例のユーザの履歴を利用した応答生成部での処理の例を示す図である。

【図13】本発明の実施例の感情判定部での「感情度」による感情のアナログ判定処理の例を示す図である。

【符号の説明】

1, 1-1, 1-2…入力部、2, 2-1, 2-2…特徴抽出部、3…感情特徴格納部、4…感情判定部、5…応答特徴格納部、6…応答生成部、7, 7-1, 7-2…出力部、8…時刻取得部、9…履歴格納部。

【図2】

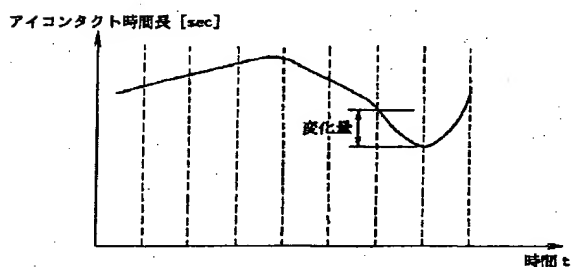


【図3】

時間 [sec]								
入力情報	ピッチ変化量[%]	+50	+10	+120	-10	-20	-85	+10

【図5】

【図4】



【図8】

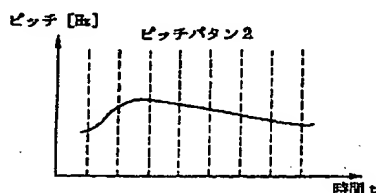
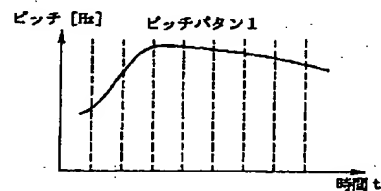
時間 [sec]								
入力情報	アイコンタクト変化量[%]	+10	-10	+45	+20	-20	-45	-10

【図6】

時間 [sec]								
入力情報	ピッチ変化量[%]	+15	+10	+30	-10	-5	-35	-5
	アイコンタクト変化量[%]	+10	-10	+45	+20	-20	-45	-10

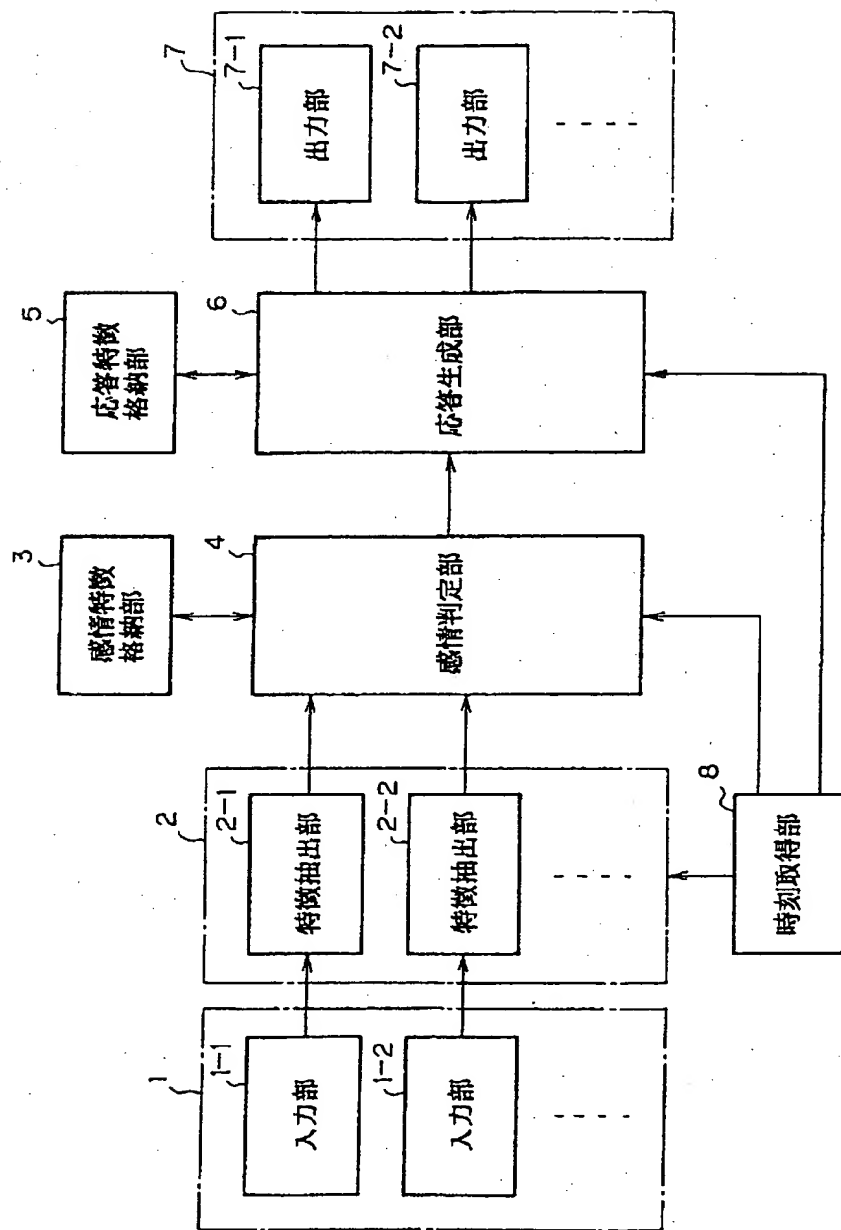


感情判定		ふつう	ふつう	楽しい	楽しい	楽しい	退屈	退屈
------	--	-----	-----	-----	-----	-----	----	----



(6)

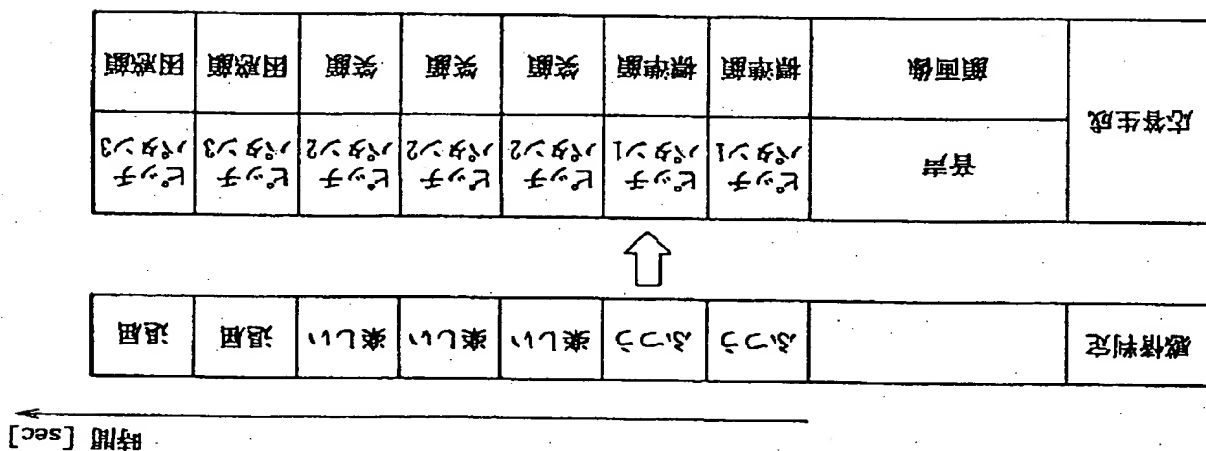
【図1】



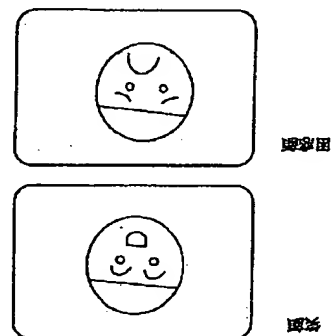


(7)

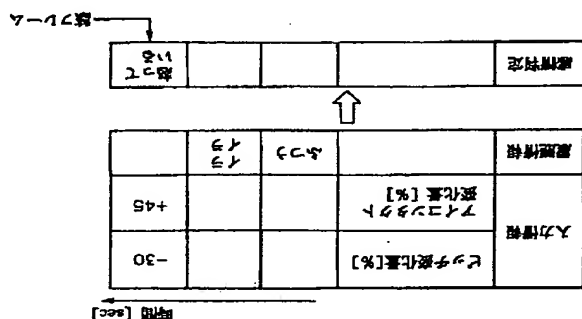
【図7】



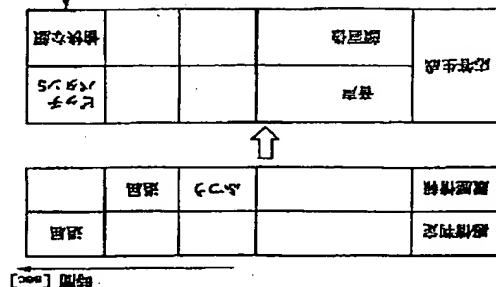
【図9】



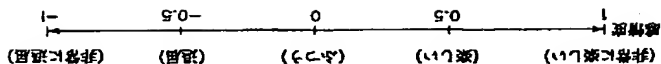
【図11】



【図12】

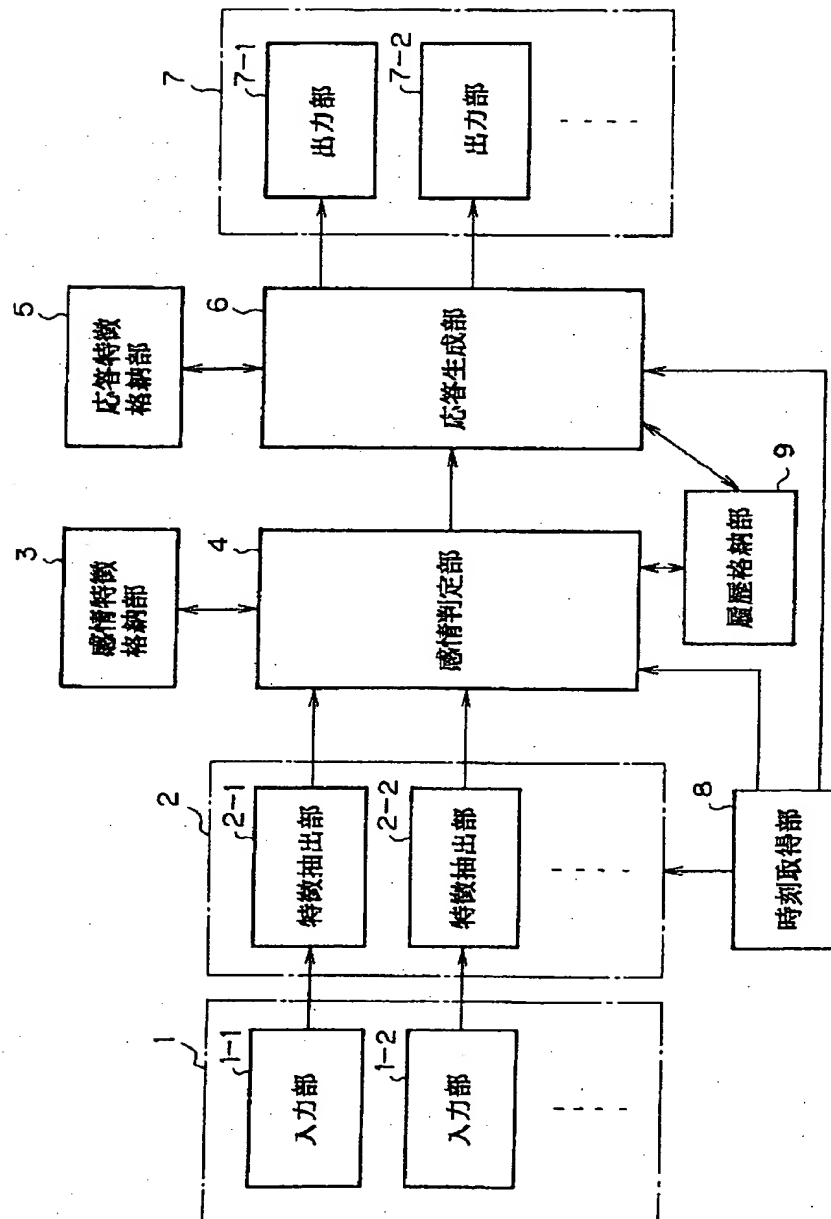


【図13】



(8)

【図10】



フロントページの続き

(51)Int. Cl. 6

G10L 3/00  
9/00

識別記号

571  
301

庁内整理番号

FI

G10L 9/00  
G06F 15/62

技術表示箇所

301A  
380